Experiment Analysis in Newspaper Topic Detection

Armelle Brun, Kamel Smaïli, Jean-Paul Haton

Experiment Analysis in Newspaper Topic Detection

Armelle BRUN, Kamel SMAILI, and Jean-Paul HATON
LORIA / INRIA-Lorraine
Campus Scientifique B.P.239
F-54500 Vandœuvre-lès-Nancy, France
{brun, smaili, jph}@loria.fr

Abstract

This paper presents several methods for topic detection on newspaper articles, using either a general vocabulary or topic-specific vocabularies. Specific vocabularies are determined manually or statistically. In both cases, we aim at finding the most representative words of a topic. Several methods have been experimented, the first one is based on perplexity, this method achieves a 100% topic identification rate, on large test corpora, when the two first propositions are taken into account. Other methods are based on statistical counts and achieve 94% of identification on smaller test corpora.
The most challenge of this work is to identify topics with only few words in order to be able, during speech recognition, to determine the best adequate language model.

1. Introduction

Current Automatic Speech Recognition (ASR) systems ar made up of two main parts, respectively devoted to acoustic modeling and language modelling. This paper deals with the second part, i.e., language models.
Stochastic language models are able to model long-range dependencies by taking into account a large history. Nevertheless, they are usually limited to model short-range dependencies: under the form of bigram or trigram models, due to the huge amount of data required to estimate reliable probabilities.
The last two words recognized by an Automatic Speech Recognition system do not contain the whole characteristics present in the text being processed, thus reducing the performance of the language model. Trigram models can be improved by inserting additional characteristics concerning words that have been recognized beyond the last two words and that are not yet integrated in the language model. Such words contain information about the current linguistic structure, the topic of the text, its style, the vocabulary is used, etc.
The texts used in our case are newspaper articles. We postulate that two articles dealing with different subjects (topics) have different behaviours, especially concerning the vocabulary. Consequently, we can consider languages used in these two articles as different, and then represent each of them by a different language model. One possibility to take into account longer ranges dependencies consists in firstly detecting the subject (topic) of the current article, and secondly integrating the topic-specific language model into the current language model, in order to modify the probabilities of predicted words. The purpose of this paper is to study several methods to identify the topic of a text. In comparision with the well known TDT DARPA-project, our aim is only to focus on the detection task, not on segmentation and tracking [12].

2. State of the Art

The first step in language model adaptation is the detection of the topic of the current document. That can be viewed as a discriminant analysis problem: given an article (document), what class(es) (topic(s)) does this article belong to? Topic detection will allow to identify the sub-vocabulary used in the article, and to more accurately determine the sub-vocabulary that will be used for the following words. The second step concerns the adaptation of the language model to the current text, thus predicting more accurately words to be recognized.

2.1. Detecting topic(s) in a text

Two points of view can be considered: one article either treats one and only one topic, or it can treat several topics. The choice of one of these two problems depends mainly on the number of topics treated (cf section 3.). If the number of topics is not large enough, only one topic is generally detected, else several topics can be searched in order (to obtain a more accurate language model).

2.1.1. Detecting one topic Topics can be detected using either the whole vocabulary of the ASR system, or a sub-vocabulary made up of topic-specific words.
Using topic-specific words consists of two steps:

- First creating, for each topic, a list of keywords. These

keywords, when occurring in one text, may contribute efficiently to topic detection: they play the role of discriminative words. The main advantage of this method is to solve the problem of noisy words (function words for example) which are not taken into account to detect topic: detection may be more accurate than by using the whole vocabulary. Creating a topic-words list for each topic is the crucial step in topic detection using sub-vocabularies. Topic-vocabularies should contain only discriminative words, as explained previously. These words are selected according to their ability to model a topic. This selectio uses a threshold related to the ability of a word to model a topic: if this threshold is too strict, detection would not be powerful, due to the shortness of the list (some texts may not contain any words of the topic-list), at the opposite, if the selection criterion is too permissive, the keyword list will be much too large (topic-vocabularies would not be very different).

- In the second step, we aim at detecting the topic for which the current text's words are the closer of this topic's keywords. Closeness can be expressed, for example, in terms of the count of topic-words recognized in the text [5], [1]. Using the whole vocabulary has the advantage, of not creating a keyword list for each topic, but may be less powerful than the previous method because of the presence (and use) of non-topic words. One of the standard methods consists in creating one language model for each topic and using these language models to detect the current topic. The resulting topic will be chosen as the one corresponding to the lowest perplexity [10],[11].

2.1.2. Detecting several topics One document can treat several topics, especially when the number of topics considered is large. Topics assigned to one article are usually the ones that best model it (using the methods presented previously) [11], [6]. Several problems are also encountered: what is the number of label-topics to assign to an article? Will this number be set *a priori* or determined dynamically?
In order to overcome this problem, Kneser in [7] does not try to detect which topics are present in a given article, he rather considers that one article contains characteristics of all treated topic. The language model is then directly adapted to the current article, according to the ability of each topic language model to represent the current article. The resulting language model is a linear combination of each topic-language model. A weight λ_i is then assigned to each topic i. $\lambda = (\lambda_1, \ldots, \lambda_k)$ is the weight vector, where k is the number of topics considered. λ corresponds to the vector that maximizes the likelihood of the part of text recognized.

Table 1: Topic labels and size of corpora available

Topic Label	**Corpus Size**
Case (CAS)	140 000
Agriculture (AGR)	200 000
Culture (CUL)	140 000
Defence (DEF)	140 000
Development (DEV)	170 000
Human Rights (HUM)	210 000
Economy (ECO)	230 000
History (HIS)	450 000

2.2. Language model adaptation

Once the topic(s) of the current document is detected (corresponding to the lowest perplexity, or the more representative keyword list), the language model is adapted to predict more accurately the next words of the text. The resulting language model can be either the language model of the topic detected [8], or the result of the combination of one language model with the general language model [7], [4], or the combination of several topic language models [3].
The number of topics taken into account in the resulting language model can nevertheless be limited: in [9], the authors prove that combining the general language model with every (weighted) topic language model does not reduce significantly the perplexity (only 3% gain) compared to combining only one topic-specific language model with the general language model.

3. How to define topics?

Before detecting topic(s) in a given article, it is first necessary to determine which topics will be considered.
The choice of topics can be implemented in two different ways:

- With the first method, articles are already grouped into classes, topics can then be directly used. The corpus we use is made up of articles from "Le Monde Diplomatique ", a French newspaper, and articles are already grouped into labelled clusters. The topics we

have chosen are the ones proposed by the newspaper. One can notice that the labels in a newspaper are only used to gather articles related to the same general idea of the topic. When one look to some articles belonging to the same cluster, one can notice that these articles could easily be put in other clusters.
We have kept only topics for which a more or less important amount of data was available. First experiments have been carried out with only 8 topics (the other topics do not contain sufficient data).
Table 1 shows which topics have been chosen and the size of each one. It can be seen that th size of the corpus are really small, and future work will also consist in collecting more data.

- The previous case is nevertheless rare enough: most of time, corpus are not yet grouped according to the topic treated in each article. Several methods have then been studied, especially [2] uses language models to classify documents. The smallest distance given by the language models between the clusters and a document allows to discover its topic. Martin in [9] also uses language models (unigrams), in order to compute what articles relate the same topic.

4. How to detect a topic?

The first method we have tested to detect the current topic consists in using language models. We make the following assumption: the label of a topic is assigned to an article, if the corresponding language model best modelizes this article.
The first problem to solve was the determination of the criterion for the evaluation of a language model. The criterion chosen is perplexity, which reflects the ability of a language model in modeling text. Perplexity is computed as follows:

$$PP(W) = \left(P(w_1) \prod_{i=2}^{N} P(w_i / w_{i-1}, \ldots, w_{i-n+1}) \right)^{-\frac{1}{n}}$$

where $W = w_1 \ldots w_N$ is the text on which the current language model will be evaluated, N is the size of the text, and finally n is the order of the language model (n-gram model). For each topic, we build a language model (based on the available training corpus). Each language model contains features of the topic it represents. To determine the topic of the current article, we compute the corresponding perplexity with each available language model. The model that provides the lower perplexity is the one that better accounts for the current text. It will then be used to label the current text.
The second method we developed consists in detecting the topic of the current article using a list of keywords for each topic. The topic we assign to a given article is the one whose keyword list is closest to the vocabulary used in this article. Several types of experiments have been conducted. The first one consists in simply counting, for each topic, the number of topic-keywords appearing in the current article. The resulting topic is the one that corresponds to the highest count of words appearing in the text. The main problem of this method is its simplicity: the same weight is assigned to each word. The second experiment consists in assigning different weights to words in a same topic-vocabulary, according to its ability to detect the current topic (for example, the word "bank" may have a greater weight than the word "international" in topic Economy. The word "bank" is actually generally associated to the Economy topic, whereas the word "international" can be related to Economy, Human Rights, or Development topics). Two different ways for computing the weight of topic-words will be studied. The first one exploits the number of topic-vocabularies in which a given word is present, and the second one is based on the probability of a given word in the topic.

5. Experiments

Each topic-corpus available is divided into two parts, for test and training. The test corpus corresponds to around 5% of the available corpus for each topic, the corpus left is assigned to the training corpus. Eight training and test corpora are then available.
A general vocabulary of 10 000 words has been extracted from the eight training corpora. Two methods for building the vocabulary can be used: the first one consists of the concatenation of the eight training corpora and then of the extraction of the 10 000 more frequent words. The second method, used to avoid the influence of the difference between the size of different corpora (for instance, the HIS corpus is 4 times larger than the CAS corpus), consists in creating a vocabulary for each topic (keeping the N more frequent words for each vocabulary). The resulting vocabulary is the concatenation of all topic-vocabularies.
Because of the small size of data, only bigram models have been processed. One bigram language model has then been built for each available topic.

5.1. Topic detection using whole vocabulary

Our first experiment consists in computing, for each test corpus, the perplexity corresponding to each topic-specific language model, using the vocabulary obtained by concatenating the eight vocabularies. The aim of the experiment is to show how close is the lowest perplexity of a topic language model and the topic label assigned by the newspaper. Table 2 contains, for each test corpus (columns), the perplexity corresponding to each topic. We can see that, for 5 test corpus out of 8, the lowest perplexity is given by the appropriate topic language model, but this is not the case for the 3 other ones for which the theoretical label is in the second position.

Table 2: Perplexity Corresponding to each Test Corpus and each Language Model

Test **Training**	**CAS**	**AGR**	**CUL**	**DEF**	**DEV**	**HUM**	**ECO**	**HIS**
CAS	172.2	226.8	226.5	274.7	212.2	206.4	245.3	259.1
AGR	247.4	**152.5**	226.3	271.6	177.6	200.6	207.9	264.8
CUL	240.3	225.4	183.6	274.9	231.9	192.5	246.4	229.9
DEF	286.8	297.8	296.6	**156.3**	284.5	274.2	319.8	308.9
DEV	259.2	195.9	230.2	263.8	**155.1**	202.5	196.5	270.7
HUM	**129.3**	217.9	243.5	236.1	209.8	169.3	256.5	208.5
ECO	228.2	201.5	**85.3**	222.1	151.5	189.2	**149.8**	235.3
HIS	180.1	200.22	193.2	198.5	185.8	**148.0**	207.5	**144.5**

We have noticed previously, that the sizes of the training corpora were very different. In order to quantify the influence of the size of the training corpora on topic detection, further experiments have been carried out. These experiments consist in reducing the size of the training corpus in order to obtain corpora containing the same number of words, a new vocabulary has then been created, based on new training corpora. This new vocabulary has been built using the first method presented above, making the assumption that using corpus of the same size, the two methods will provide equivalent vocabularies.
The results obtained with the new corpora are comparable to the ones obtained in previous experiments. We can thus conclude that corpus size has no influence on topic detection performance.

5.2. Results Analysis

CAS, CUL and DEV test corpora have been studied to search the reason for the failure of the detection for these topics.
CAS corpus is composed of 4 articles, topic detection has been processed on each of these four articles, the lower perplexity for each one should correspond to CAS topic. One of the 4 articles has the lower perplexity corresponding to the HUM topic, the second rank corresponds even so to CAS topic. Surprisingly, this article is labelled CAS and HUM in the newspaper. The same tests have been performed for Culture and Development topics where the same kind of problem has been detected and explained. We can conclude that the results obtained using the method presented here are consistent, our method seems to perform efficiently.
We will now study the adequacy of the general vocabulary (based on training corpora containing the same count of words) to each topic. The general vocabulary has been created using the concatenation of every topic corpora, keeping the 10 000 more frequent words. We first evaluate the rate of unknown words in each topic in order to identify which topics are not sufficiently represented by the general vocabulary, and then we study the coverage of each topic. Coverage is evaluated as follows:

$$cov_{T_i} = \frac{N(V \cap V(T_i))}{N(V)}$$

where V is the vocabulary, $V(T_i)$ is the list of distinct words found in the training corpus topic T_i, $N(V \cap V(T_i))$ is the count of words of the general vocabulary that are also present in the topic training corpus, and $N(V)$ is the size of the general vocabulary. Results of this study are presented in Table 3.

We can remark that HIS, CAS and CUL topics have the highest rate of unknown words, this means that these three topics are underrepresented. The contribution of their vocabularies to the general vocabulary should be increased. We can also notice that the Defense topic has a coverage of 66.6%, and a percentage of unknown words of 4.67% (the

Table 3: Adequacy of the General Vocabulary to each Topic

Topic	**% Unknown Words**	**coverage**
Cases	7.09	75.2
Agriculture	5.80	72.1
Culture	7.95	76.1
Defence	4.67	66.6
Development	5.16	72.2
Human Rights	6.57	73.9
Economy	5.62	73.9
History	8.31	76.7

lowest percentage). We can conclude that this topic seems to require less words to be represented. Similar conclusions can be drawn for the Development topic. These remarks make it possible to conclude that some topics need more words to be represented than others. The same table should be processed using a vocabulary based on the concatenation of topic-vocabularies, in order to make more reliable conclusions about Defense and Development topics.
Our methods for constructing the general vocabulary may be improved. A solution could be to create topic-vocabularies that correspond to a fixed coverage (for example 80%), the resulting general vocabulary being the concatenation of each topic-vocabulary.

5.3. Topic Detection in Speech Recognition

In our case, topic detection is used in the domain of automatic speech recognition, in order to adapt the language model to the text being processed. Topics must then be detected as soon as possible, when a minimum of words have been pronounced. Experiments have thus been conducted to study the performance of our methods on few words. Topic detection has then been processed using a test corpus of 10, 20 and until 200 words. Test corpus contains around three hundred words. Topic detection performance is presented in Figure 1. Topic detection is powerful on test corpus with size of around one article, but it is less powerful on smaller test corpus. In the case of a set of sentences of about 200 words long, around 54% of the test corpus are correctly detected in first rank, 77% are detected in the 2 first ranks and 92% in the first 3 ranks. Most of test sentences treat several topics, that is why the topic given by the newspaper is not always correctly identified but is generally in the first 3 ranks. The other topics ranked in the 3 first places are not surprising. These results will be discussed in Section 6.
Topic detection using texts containing less than 60 words is not reliable: in such cases curves seem to indicate a random behavior (data not shown).

We can notice that the increase of the number of words taken into account (from 60 to 200) results in an increase of 15% in topic detection performance. Little benefit in topic detection seems to be expected from increasing the number of words over 160.

5.4. Topic detection using sub-vocabularies

Previous experiments have been conducted on a general vocabulary. It is interesting to check if words that are not topic words, do not decrease the performance. Experiments have been conducted to detect topic using topic-vocabularies. Three types of vocabularies have been used to study our topic-detection methods:

- The first method for creating topic vocabularies consists in constructing topic-lists containing the most frequent words in each training corpus. The major problem with this method is that overlap rate exceeds 60%, due to function words, which are in every topic-vocabulary.
- The second method consists in creating topic-vocabularies composed of the more discriminant words. Indeed, we remarked that some words are in every topic-vocabulary (the word *justice*, for instance). Other words are also in almost every vocabulary (7 out of 8), these words are not useful for topic identification (discrimination). We have carried out experiments using the number of topic-vocabularies in which a word appears as discrimination criterion: every word appearing in more than half of vocabularies was removed.
- The last vocabulary tested has been built manually: we created for each topic, a vocabulary composed of words appearing more than 5 times in the training corpus, we then remove manually from these vocabularies every word not representative of the topic, thus obtaining topic-vocabularies containing around two hundred words. Table 4 summarizes the 3 types of vocabularies used in our topic detection methods presented in this paper.

In the following, we will use the above vocabularies with three different methods (M_1,M_2 ,M_3) in order to identify the method and vocabulary which yield the best results. All the experiments described below have been carrried out with a test corpus of 200 hundred words for each topic, in

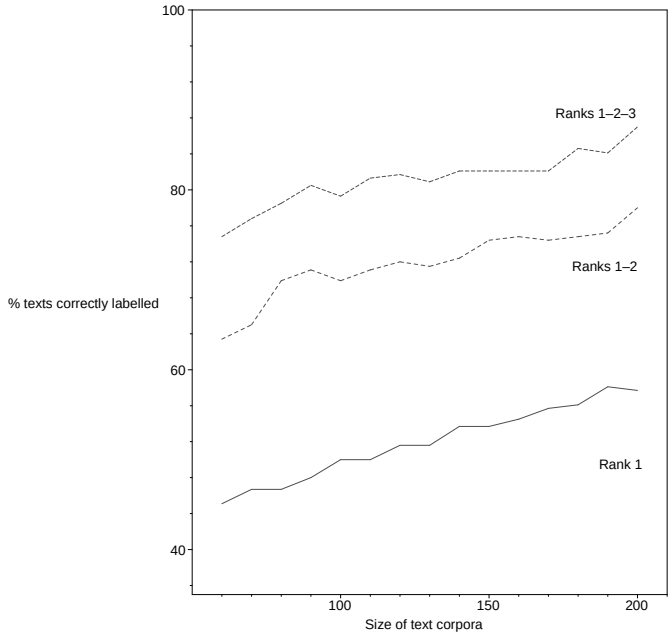

Figure 1: Topic detection performances in accordance to the number of words recognized

Table 4: M_2 topic identification performances

Vocabulary	**Composition of the Vocabulary**
V_1	Words appearing more than 5 times in the training corpora
V_2	Words appearing in at most 4 of the preceding topic-vocabularies
V_3	Vocabulary built manually

order to be in the situation of language model adaptation in speech recognition.

5.5. A basic counting-topic-words method (M_1):

The first method (basic) we have used to detect a topic consists in simply counting the number of topic-words occurring in the test corpus, the topic corresponding to the larger number is considered as the topic of the test corpus. Topic detection performances have been studied, using the three vocabularies. The results are presented in Table 5.

Topic detection using vocabulary V_1 is as powerful as the detection using perplexity: for each test corpus (same as in Table 2), the theoretical topic (given by the newspaper) is the one corresponding to the higher number of topic-words detected in 54% of the cases. This percentage increases when the 2 first ranks are taken into account (79%), and reaches 87.5% with the 3 first ranks (less powerful than using a general vocabulary (Figure 1)). It can be notified that the score obtained for the theoretical topic is often equal or only slightly larger than those given of other topics. Also, the difference between the scores of each topic is not significantly different (for a phrase of 80 words, around 65 topic-words are detected and the difference between the greater score and the worse is around 5 words), reliable conclusions can then not be made.

Further experiments consisting in removing function words from vocabularies have been conducted. The number of topics correctly detected has not been improved and the two points just noticed still remain.

Vocabulary V_2 has been used to detect topic of texts. The detection is more powerful than by using a vocabulary composed of words appearing more than n times: 54% of the 200-words-phrases are recognized in rank one, 77% in the second rank, and 86% in the three first ranks (which is equivalent to the results provided with the first type of vocabulary). The difference between scores is larger, but still not reliable.

Topic detection has then been done using vocabulary V_3 on a set of sentences of 200 words, 49% of the theoretical topics are detected in first rank, 83% in the two first ranks and 94% in the three first ranks. This experiment has been done in order to determine the upper bound of the ability of topic detection of our method. We can notice that results obtained using the manually built vocabulary (V_3) are better than the ones obtained using vocabulary V_1, as expected.

A comparison between vocabulary V_1 and vocabulary V_2 in terms of topic detection performance is presented in Figure 2.

Topic detection, especially in first rank does not exceed 54%, the reason is certainly the simplicity of our method (simple count). This performance may be improved using more sophisticated methods. A second method fot topic detection based on weighted counting willnow be presented:

5.6. A weighted count topic identification (M_2):

The method presented here consists in giving different weights to topic-words, as explained in Section 4.. One method consists in giving a weight conversely proportional to the number of topic-vocabularies in which this word appears (for example, a word appearing in 4 topic-vocabularies will have a weight of $\upsilon(w_i) = \frac{1}{5}$). The score of a phrase (S) computed for a topic j is given by:

$$T_j = \frac{1}{kU^U} \prod_{i=1}^{k} \phi_j(w_i)$$

where $U = (k + 1 - N_k(V_j))$ and $\phi(w_i) = \upsilon(w_i)\frac{1}{N+1}$ (with $\sum_{j=1}^{N_T} \sum_{i=1}^{N_j} \phi_j(w_i) = 1$)

k being the number of words in S, the word at position i, j the topic under test and $N_k(V_j)$ the number of words of vocabulary V_j encountered in the k first words of S. N corresponds to the cumulative distinct number of words for T topic-vocabularies. ϕ depends on the inverse of $N + 1$. We added one in order to be able to give a weight to off-topic words. U is a term which allows to reduce the score of T_j in accordance with the number of off-topic words.

Topic detection results obtained with this method are presented in Table 6.

When can notice that results obtained using vocabulary V_3 are similar to the ones using method M_1 . The reason is certainly the overlap rate between vocabularies V_3 that approaches zero. This method has to be improved by better estimating the off-topic words weight. Moreover, in both methods presented here, vocabularies built manually provide the best results (especially concerning the 3 first ranks), assuming that results obtained with method M_1 are biased (see section 5.5.). The Evolution of topic detection performances using vocabulary V_1 is presented in Figure 3.

5.7. A unigram topic identification (M_3):

The last method, as for it, gives a weight to a word based on its topic-unigram probability. For example, to study if a text treats the CUL topic, the weight of word w_i appearing in this text will correspond to the unigram probability of this word in CUL topic. At the opposite, if we study DEF topic, the weight of this word will correspond to its unigram probability in this topic, which is different from the one in CUL. Results obtained are presented in Table 7:

We remark that this method is the one that provides the worst results. The reason is certainly the small size of training data which does not allow to compute reliable probabilities. Vocabulary V_1 seems to be the one that provides the worst results, the reason is the overlaprate between each vocabulary, which exceeds 60%. Vocabulary V_3 always provides the best results. The main problem concerning this vocabulary is that it contains only words that are representative of the manually selected topic. The size of these

Table 5: M_1 topic identification performances

	Rank 1 (%)	**Ranks 1-2 (%)**	**Ranks 1-2-3 (%)**
V_1	54	79	87
V_2	54	77	86
V_3	49	83	94

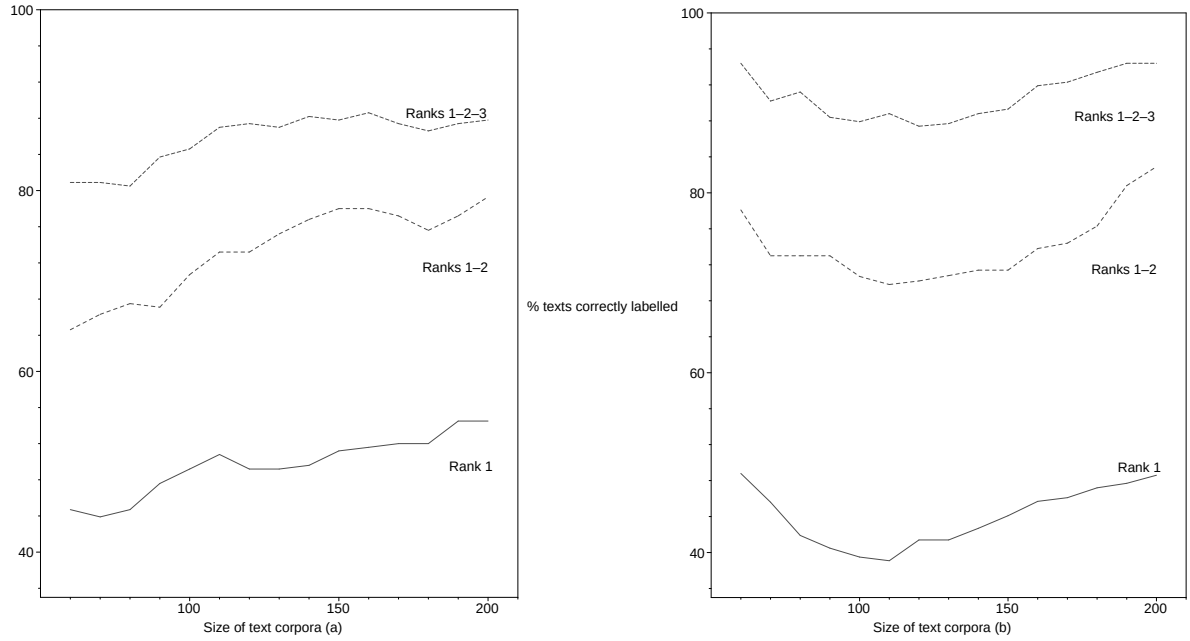

Figure 2: Variation of topic identification rate with the number of words for vocabularies V_1 (a) and V_3 (b)

Table 6: M_2 topic identification performances

	Rank 1 (%)	**Ranks 1-2 (%)**	**Ranks 1-2-3 (%)**
V_1	52	77	89
V_2	52	72.5	83
V_3	49.5	83	93.5

Table 7: M_3 topic identification performances

	Rank 1 (%)	**Ranks 1-2 (%)**	**Ranks 1-2-3 (%)**
V_1	13	34	41
V_2	56	76	85
V_3	57	80	84.5

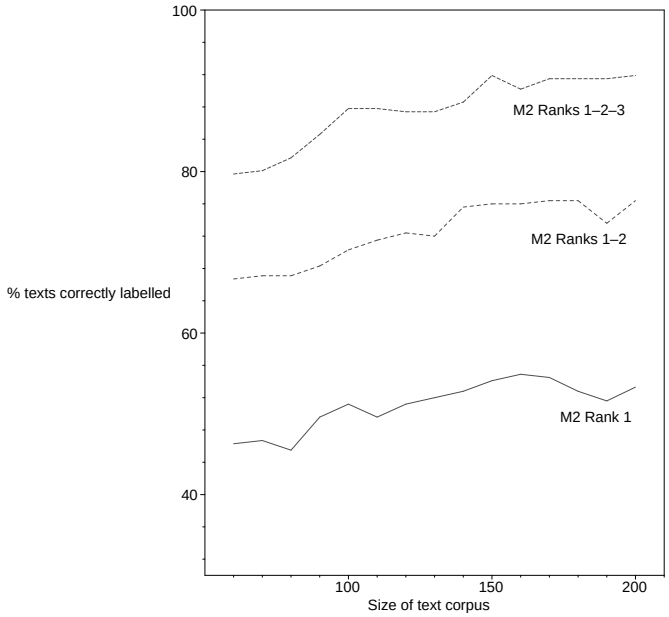

Figure 3: Evolution of topic detection with the number of words, using vocabulary V_1 and method M_2

vocabularies is very small, which is one of the two weaknesses of this type of method, the other one concerns the way of building this vocabulary. Building manually 8 topic-vocabularies is feasable, but that will no longer be possible, when the number of topics increases.

6. Discussion and Conclusion

In this paper, several methods and vocabularies have been developped, to cope with the problem os topic-identification in speech recognition. The basic methods developed gave promissing results. The main problem we encoutered is the fact of testing these methods with only few words in order to be able to adapt language models in speech recognition (contrary to many other works that use whole texts to detect one topic). If we consider the three best propositions obtained, we reach a rate of 94% of good topic-identification. As noted before, some articles belong two or three topics. That is why we presented our results in terms of the three best propositions. For the adaptation, we will probably have to combine the corresponding language models. The topic of the text (at the semantic level) does not always correspond to the vocabulary used in this article: let consider one article treating agriculture: the beginning of the article could be ”summer is the season of crops ”, then can be a comma, the article can then relate life fits of farmers during the last century. The main topic of the article is evidently Agriculture, the apposition treating Human Rights and History. Topic detection can then be viewed from two points of view:

- one can either search at detecting the main topic of the given article, we get placed at the semantic level
- or one can get interested at the lexical level, we search at detecting every topic treated in the text, which is our case.

Indeed, we aim at detecting topic in order to adapt the language model used in the ASR system, more especially, we search at predicting the future words to appear. In the case of the article presented previously, words to appear can belong obviously to Agriculture topic, but as well to Human Rights and History topics. The three topics must be detected by our system.
One major problem does also appear: articles used in training and test corpora are labelled mainly using the first point of view presented here, which doesn't correspond to our view of topic detection. We can then conclude that training corpora used in this article are not perfectly representative of our topics.

7. References

[1] B. Bigi, R. De Mori, and M. El Bèze. Detecting topic shifts using a cache memory. In *Proceedings of International Conference on Spoken Language Processing*, 1998.

[2] Beth. A. Carlson. Unsupervised topic clustering of switchboard speech messages. In *Proceedings of International Conference on Spoken Language Processing*, 1996.

[3] P. R. Clarkson and A. J. Robinson. Language model adaptation using mixtures and an exponentially decaying cache. In *Proceedings of International Conference on Spoken Language Processing*, 1997.

[4] S. Della Pietra, V. Della Pietra, R. L. Mercer, and S. Roukos. Adaptive language modelling using minimum discriminant estimation. In *Proceedings of International Conference on Spoken Language Processing*, 1992.

[5] Mc Donough, K Ng, and P. JeanRenaud. Approaches to topic identification on the switchboard corpus. In *Proceedings of International Conference on Spoken Language Processing*, 1994.

[6] T. Imai, R. Schwartz, F. Kubala, and L. Nguyen. Improved topic discrimination od broadcast news using a model of multiple simultaneous topics. In *Proceedings of International Conference on Spoken Language Processing*, 1997.

[7] Reinhard Kneser and Volker Steinbiss. On the dynamic adaptation os stochastic language models. In *Proceedings of International Conference on Spoken Language Processing*, 1993.

[8] R. Lau. *Adaptive Statistical Language Modelling*. PhD thesis, Carnegie Mellon University, 1994.

[9] Sven C Martin, Jorg Liermann, and Hermann Ney. Adaptive topic-dependent language modeliing using word-based varigrams. In *Proceedings of Eurospeech Conference*, 1997.

[10] S. Matsunaga, T. Yamada, and K. Shikano. Task adaptation in stochastic language models for continuous speech recognition. In *Proceedings of International Conference on Spoken Language Processing*, 1994.

[11] R. Schwartz, T. Imai, F. Kubala, and L. Nguyen. A maximum likelihood model for topic classification of broadcast news. In *Proceedings of Eurospeech Conference*, 1997.

[12] http://www.itl.nist.gov/iaui/894.01/tdt98/tdt98.htm, 1998.